Appunti di Probabilità e Statistica

a.a. 2014/2015 C.d.L. Informatica – Bioinformatica
I. Oliva

Lezione 5

1 Momenti di v.a.

Vogliamo associare alle v.a. delle grandezze deterministiche, che ne diano informazioni qualitative e quantitative.

Media Se X è una v.a. discreta con supporto $S = \{k : k \in I\}$, $I \subset \mathbb{Z}$, e con densità $P(X = k)$ e se supponiamo che $\sum_{k \in S} |k| P(X = k) < \infty$, si chiama *media* di X la quantità

$$E(X) := \sum_{k \in S} k P(X = k) \; .$$

Esempio 1.1. *X : guadagno netto puntando un euro sul rosso alla roulette onesta, Y : guadagno netto puntando un euro su T nel lancio di una moneta equa.*

X, Y sono v.a. discrete, entrambe con supporto ± 1. Inoltre, $P(X = 1) = 18/37$, $P(X = -1) = 19/37, P(Y = 1) = P(Y = -1) = 1/2$. Allora, avremo:

$$E(X) = 1 \cdot \frac{18}{37} + (-1) \cdot \frac{19}{37} = -\frac{1}{37}$$
$$E(Y) = 1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0$$

Se X è una v.a. continua con densità f_x e se supponiamo che $\int_{\mathbb{R}} |x| f_X(x) \mathrm{d}x < \infty$, si chiama *media* di X la quantità

$$E(X) := \int_{\mathbb{R}} x f_X(x) \mathrm{d}x \ .$$

Proprietà della media

1. (*coerenza*) Se $P(X = c) = 1$, allora $E(X) = c$. Infatti:
   se $P(X = c) = 1$, allora, per definizione di media, $E(X) = c \cdot P(X = c) = c$.
2. (*linearità*) Se $E(X) < \infty$ e a, b sono due costanti, allora $E(aX + b) = aE(X) + b$. Infatti:
   se X è v.a. discreta, allora
   $$\begin{aligned} E(aX+b) &= \sum_{k \in I} (ak+b) P(X=k) = a \sum_{k \in I} kP(X=k) \\ &\quad + b \sum_{k \in I} P(X=k) = aE(X) + b; . \end{aligned}$$
   Se X è v.a. continua, allora
   $$\begin{aligned} E(aX+b) &= \int_{\mathbb{R}} (ax+b) f_X(x) \mathrm{d}x \\ &= a \int_{\mathbb{R}} x f_X(x) \mathrm{d}x + b \int_{\mathbb{R}} f_X(x) \mathrm{d}x = aE(X) + b; . \end{aligned}$$
3. (*internalità*) Se $a, b \in \mathbb{R}$ sono tali che $P(a \leq X \leq b) = 1$, allora $a \leq E(X) \leq b$. Infatti:
   se $P(a \leq X \leq b) = 1$, allora, per definizione di media, esiste $E(X)$. In particolare, $P(X - a) \geq 0$, dunque
   $$0 \leq E(X - a) = E(X) - E(a) = E(X) - a \ .$$
   Analogamente, $P(b - X) \geq 0$, dunque
   $$0 \leq E(b - X) = E(b) - E(X) = b - E(X) \ .$$

La media si chiama anche *momento primo,* in quanto si possono definire i *momenti di ordine* i, per ogni $i \geq 1$:

$$E(X^i) = \sum_{k \in I} k^i P(X = k) \ .$$

Varianza Se X è una v.a. (discreta o continua), per cui esiste $E(X)$ e $E((X-E(X))^2)$, si chiama *varianza* di X la quantità

$$Var(X) = E((X-E(X))^2) \ .$$

Esempio 1.2. *Y : guadagno netto puntando 1 € su T nel lancio di una moneta equa, Z : guadagno netto puntando 1000 € su T nel lancio di una moneta equa. Avremo:*

$$P(Y=1) = P(Y=-1) = 1/2, \ P(Y=1000) = P(Y=-1000) = 1/2 \ .$$

Inoltre, $E(Y) = E(Z) = 0$. Per quel che riguarda la varianza:

$$\begin{aligned}
Var(Y) &= E((Y-E(Y))^2) = E(Y^2) \\
&= \sum_{k \in \{-1,1\}} k^2 P(X=k) = (-1)^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{2} = 1 \\
Var(Z) &= E((Z-E(Z))^2) = E(Z^2) \\
&= \sum_{k \in \{-10^3,10^3\}} k^2 P(X=k) = (-1000)^2 \cdot \frac{1}{2} + 1000^2 \cdot \frac{1}{2} = 10^6
\end{aligned}$$

Proprietà della varianza

1. $Var(X) = 0$ se e solo se $P(X=c) = 1$. Infatti:
   se $P(X=c) = 1$, allora $E(X) = c$ e $Var(X) = E((c-c)^2) = 0$. Viceversa, se $Var(X) = 0$, essendo $P((X-E(X))^2 \geq 0) = 1$, allora $P((X-E(X))^2 = 0) = 1$, che si verifica se e solo se $P(X = E(X)) = 1$.
2. Se X ammette varianza e $a, b \in \mathbb{R}$, allora $Var(aX+b) = a^2 Var(X)$. Infatti:

   $$\begin{aligned}
   Var(aX+b) &= E((aX+b-E(aX+b))^2) \\
   &= E((aX+b-aE(X)-b))^2) = E((aX-aE(X)))^2) \\
   &= a^2 E((X-E(X))^2) = a^2 Var(X).
   \end{aligned}$$

3. Se X ammette varianza, allora X^2 ammette media e $Var(X) = E(X^2) - (E(X))^2$. Infatti:

Se X ammette varianza, allora

$$\begin{aligned} E(X^2) &= E((X - E(X) + E(X))^2) \leq E((X - E(X))^2 + E(X))^2) \\ &\leq 2E((X - E(X))^2) + 2E(X)^2 \\ &= Var(X) + 2E(X)^2 < \infty \end{aligned}$$

Questo garantisce che il momemento secondo esiste. Inoltre:

$$\begin{aligned} Var(X) &= E((X - E(X))^2) = E(X^2 - 2XE(X) + E(X)^2) \\ &= E(X^2) - E(X)^2. \end{aligned}$$

Esercizio 1.1. *Sia* $X \sim Bin(n, p)$. *Allora* $E(X) = np$ *e* $Var(X) = np(1 - p)$.

Esercizio 1.2. *Sia* $X \sim Po(\lambda)$. *Allora* $E(X) = \lambda$ *e* $Var(X) = \lambda$.

Esercizio 1.3. *Sia* $X \sim \mathcal{U}(0, 1)$. *Allora* $E(X) = \frac{1}{2}$ *e* $Var(X) = \frac{1}{12}$. *In generale, se* $X \sim \mathcal{U}(a, b)$, *allora* $E(X) = \frac{a+b}{2}$ *e* $Var(X) = \frac{(a+b)^2}{12}$.

Esercizio 1.4. *Sia* $X \sim Exp(\mu)$. *Allora* $E(X) = \frac{1}{\lambda}$ *e* $Var(X) = \frac{1}{\lambda^2}$.

Esercizio 1.5. *Sia* $X \sim \mathscr{N}(0, 1)$. *Allora* $E(X) = 0$ *e* $Var(X) = 1$. *In generale, se* $X \sim \mathscr{N}(\mu, \sigma^2)$, *allora* $E(X) = \mu$ *e* $Var(X) = \sigma^2$.

2 Teoremi classici per le variabili aleatorie

Esistono alcuni importanti risultati che permettono di determinare una approssimazione della probabilità, quando le variabili aleatorie coinvolte non hanno distribuzione nota.

Disuguaglianza di Markov Sia X una v.a. non negativa, allora, per qualunque valore $a > 0$, si ha

$$P(X \leq a) \leq \frac{E(X)}{a}\ .$$

Proof. Verifichiamo la disuguaglianza nel caso di v.a. continue (il caso discreto risulta analogo). Sia X una v.a. con densità f, allora

$$\begin{aligned} E(X) &= \int_{\mathbb{R}} xf(x)\mathrm{d}x = \int_0^{\infty} xf(x)\mathrm{d}x = \int_0^a xf(x)\mathrm{d}x + \int_a^{\infty} xf(x)\mathrm{d}x \\ &\geq \int_a^{\infty} xf(x)\mathrm{d}x \geq \int_a^{\infty} af(x)\mathrm{d}x = a\int_a^{\infty} f(x)\mathrm{d}x \\ &= aP(X \geq a). \end{aligned}$$

$\square$

Esempio 2.1. *Una moneta non equa, in cui la probabilità che esca testa è* $1/10$, *viene lanciata* 200 *volte. Determinare una stima della probabilità che, su* 200 *lanci, esca testa almeno* 120 *volte.*

Sia X *la v.a. che conta il numero di teste su* $n = 200$ *lanci, allora* $X \sim Bin(200, 1/10)$. *Si tratta di calcolare, allora,* $P(X \geq 120)$, *ma è un numero troppo grande da calcolare "a mano"! Usiamo, dunque, la Disuguaglianza di Markov con* $a = 120$ *e* $E(X) = np = 200 \cdot \frac{1}{10} = 20$ *e avremo*

$$P(X \geq 120) \leq \frac{20}{120} = \frac{1}{6} \, .$$

Disuguaglianza di Chebychev Sia X una v.a. con media $E(X)$ e varianza $Var(X)$. Allora, dato un qualunque valore $k > 0$, vale la seguente relazione

$$P(|X - E(X)| \geq k) \leq \frac{Var(X)}{k^2} \, .$$

Proof. Si consideri l'evento $A = \{|X - E(X)| \geq k\}$. I valori di X e k che soddisfano la disuguaglianza dell'evento A sono gli stessi per cui si verifica l'evento $B = \{(X - E(X))^2 \geq k^2\}$, quindi la probabilità che si verifichi l'evento A o l'evento B è la stessa. Inoltre, la variabile casuale $(X - E(X))^2$ è non negativa (essendo un quadrato), dunque si può applicare la disuguaglianza di Markov, con $a = k^2$, quindi

$$P(|X-E(X)| \geq k) = P((X-E(X))^2 \geq k^2) \leq \frac{E[(X - E(X))^2]}{k^2} = \frac{Var(X)}{k^2} \, .$$

□

Esempio 2.2. *Il numero di automobili prodotte da una fabbrica in una settimana si distribuisce secondo una v.a.* X *con media pari a* 50.

Qual è la probabilità che la produzione superi occasionalmente le 75 *auto? E qual è la probabilità che la produzione sia compresa tra* 40 *e* 60 *pezzi, sapendo che la varianza della distribuzione è pari a* 25?

Usiamo la Disuguaglianza di Markov per risolvere il primo punto:

$$P(X \geq 75) \leq \frac{50}{75} = 0.67$$

Per quel che riguarda la seconda domanda, si applica la Disuguaglianza di Chebychev con $k = 10$

$$\begin{aligned} P(40 \leq X \leq 60) = 1 - P(X \leq 40, X \geq 60) &= 1 - P(|X - 50| \geq 10) \\ &\leq \frac{25}{100} = 0.75 \, . \end{aligned}$$

Legge dei Grandi Numeri Sia $\{X_n\}$ una successione di v.a.i.i.d. che ammettono momento secondo e sia $S_n := \frac{1}{n}\sum_{j=1}^n X_j$. Allora, per ogni $k > 0$, si ha

$$\lim_{n\to\infty} P(|S_n - E(X)| > k) = 0 \; .$$

A parole, la media dei campionamenti converge, in senso stocastico, al valore atteso della popolazione, quando la dimensione del campione aumenta.

Proof. Siano m, s^2 la media e la varianza di X_j, rispettivamente. Essendo le v.a. i.d., allora

$$E(S_n) = E\left(\frac{1}{n}\sum_{j=1}^n X_j\right) = \frac{n \cdot m}{n} = m \; .$$

Inoltre, essendo le v.a. indipendenti, si ha:

$$Var(S_n) = Var\left(\frac{1}{n}\sum_{j=1}^n X_j\right) = \frac{1}{n^2}\sum_{j=1}^n s^2 = \frac{s^2}{n} \; .$$

Applichiamo la Disuguaglianza di Chebychev e avremo

$$P(|S_n - m| > k) \leq \frac{s^2}{nk^2} \to 0 \; .$$

□

Teorema del Limite Centrale Sia $\{X_n\}$ una successione i.i.d. di v.a. che ammettono momento secondo e con varianza non nulla. Posto $m = E(X_n), s^2 = Var(X_n), S_n = \frac{1}{n}\sum_{j=1}^n X_j$ e $Z_n = \frac{S_n - m}{s}\sqrt{n}$, allora, per ogni $x \in \mathbb{R}$, si ha

$$\lim_{n\to\infty} P(Z_n \leq x) = P(Z \leq x) \; ,$$

dove $Z \sim \mathscr{N}(0,1)$.

Vediamone l'applicazione (non la dimostrazione) con un esempio.

Esempio 2.3. *Si lancia n volte un dado equo.*

1. *Se $n = 1000$, qual è la probabilità che il punteggio totale sia minore o uguale a 3400?*

2. *Quanto grande deve essere n perché, con probabilità maggiore o uguale del 99%, il punteggio totale sia almeno $3.3n$?*
3. *Quanto grande deve essere n perché, con probabilità maggiore o uguale del 99%, il punteggio totale sia almeno 500?*

Risolviamo i diversi quesiti. L'idea generale è considerare l'i-simo lancio del dado ed indicare con X_i il valore ottenuto, quindi prendere in esame la quantità

$$Z_n = \frac{S_n - m}{s}\sqrt{n}\ .$$

Infine, sostituire Z_n con Z per ottenere una approssimazione esplicita, ma calcolabile.

1. *Si ha $X_i \in \{1,2,3,4,5,6\}$ e $P(X_i = k) = 1/6$, per ogni $k = 1,\dots,6$. Dunque, $E(X) = 3.5$ e $Var(X) = 2.917$. Allora,*

$$\begin{aligned} P(S_{1000} \le 3400) &= P(X_1 + \cdots + X_{1000} \le 3.4) \\ &= P\left(\frac{S_n - m}{s}\sqrt{n} \le \frac{3.4 - m}{s}\sqrt{n}\right) = P(Z_n \le -1.85) \\ &\cong P(Z \le -1.85) = \Phi(-1.85) = 1 - \Phi(1.85) = 0.0322. \end{aligned}$$

2. *Determiniamo n tale che $P(X_1 + \cdots + X_n \ge 3.3n) \ge 0.99$. Quindi,*

$$\begin{aligned} 0.99 \le P(S_n \ge 3.3n) &= P\left(Z_n \ge \frac{3.3 - m}{s}\sqrt{n}\right) \\ &= P(Z_n \ge -0.117n) \cong P(Z \ge -0.117n) = \Phi(0.117n) \\ \Rightarrow 0.117\sqrt{n} \ge 2.33 \ \Rightarrow\ \sqrt{n} &\ge \frac{2.33}{0.117} = 19.91 \ \Rightarrow\ n \ge 396. \end{aligned}$$

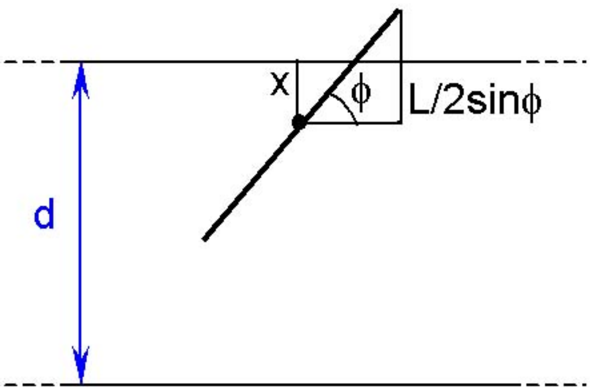
3. *Determiniamo* n *tale che* $P(X_1 + \cdots + X_n \geq 500) \geq 0.99$. *Quindi,*

$$
\begin{aligned}
0.99 \leq P(S_n \geq 500) = P & \left(Z_n \geq \underbrace{\frac{\frac{500}{n} - m}{s}\sqrt{n}}_{Q} \right) \\
& \cong P(Z \geq Q) \\
& \Rightarrow \frac{-\frac{500}{n} - m}{s}\sqrt{n} \geq 2.33 \\
& \Rightarrow -\frac{500}{n}\sqrt{n} + 3.5\sqrt{n} \geq 3.97 \\
& \Rightarrow 3.5n - 3.97\sqrt{n}\sqrt{n} - 500 \geq 0 \\
& \Rightarrow \sqrt{n} = 12.53 \;\Rightarrow\; n \geq 158.
\end{aligned}
$$

La Legge dei Grandi Numeri ed il Teorema del limite centrale hanno un vasto bacino di applicazioni:

- *Approssimazione di* π.

  Un ago di lunghezza L viene gettato a caso in un pavimento, formato da linee parallele, distanti d l'una dall'altra e supponiamo $L < d$. Ci sono due alternative: o l'ago cade esattamente tra due righe, oppure ne interseca una. Qual è la probabilità che si verifichi la seconda eventualità?

  Il setting matematico è il seguente: sia α l'angolo formato dall'ago e da una delle righe del pavimento e sia x la distanza tra il centro dell'ago e la riga più vicina. Supponiamo che φ, x siano due v.a.i.i.d. (equiprobabili). In particolare,

$$AB = L,\; OK =: x,\; 0 \leq x \leq d/2,\; B\hat{O}H =: \varphi, 0 < \varphi\ \pi\ .$$

L'ago cade su una riga se

$$x < BH = \frac{L}{2}\sin(\varphi) \ .$$

Dunque, la probabilità che cerchiamo si calcola come rapporto tra casi favorevoli e casi possibili, dove i "casi possibili" sono tutti i punti del piano, compresi tra $[0, d/2]$ e $[0, \pi]$, mentre i "casi possibili" sono tutti i punti del piano compresi tra l'equazione che descrive x e $[0, \pi]$, ossia

$$P_x = \frac{\int_0^\pi \frac{L}{2}\sin(\varphi)\mathrm{d}\varphi}{\frac{\pi d}{2}} = \frac{\frac{L}{2}[-\cos(\varphi)|_0^\pi]}{\frac{\pi d}{2}} = \frac{2L}{\pi d} \ .$$

Quanto vale questo numero? Supponiamo di effettuare N (grande) lanci e sia k il numero di volte in cui l'ago non interseca una riga, allora

$$P_x = \lim_{N\to\infty}\frac{k}{N} = \frac{2L}{\pi d} \ \Rightarrow \ \pi \approx \frac{2LN}{kd} \ ,$$

che rappresenta una stima per π.

Discorso analogo vale se $L \geq d$: ponendo $x = \frac{L}{d}$, la probabilità, cercata è

$$P_x = \frac{2}{\pi}\left(x - \sqrt{x^2-1} + \arcsin(x)\right) \ .$$

- *Metodo Monte Carlo.*

  È un metodo introdotto per calcolare la media di v.a. quando non si possono usare le tecniche dirette.

  Se X è una v.a. e se f è una funzione di X, supponiamo di poter simulare n valori di X, diciamo $x_1, \ldots, x_n$. Allora,

$$E[f(X)] \approx \frac{1}{n}\sum_{i=1}^n f(x_i) \ .$$

  Poichè si tratta di approssimazioni, dobbiamo essere sicuri che il valore ottenuto sia "buono": la Legge dei Grandi Numeri ci assicura che il valore approssimato converge al valore vero, mentre il Teorema del limite centrale ci dice che

$$\frac{1}{n}\sum_{i=1}^n f(x_i) \sim \mathscr{N}\left(E(f(X)), \frac{1}{n}Var(f(X))\right) \ .$$

Quindi, l'approssimazione ottenuta si discosta dal valore vero di una quantità dell'ordine di $\frac{1}{\sqrt{n}}$, allora, maggiore sarà n, migliore sarà la nostra approssimazione.

Il metodo Monte Carlo è particolarmente adatto per approssimare il valore di integrali definiti:

$$\int_a^b f(x)\mathrm{d}x, \quad \text{essendo } f : [a, b] \to [c, d] \, .$$

Procediamo come segue:

- sia $A := (b - a) \cdot (d - c)$ e sia $j := 0$ (contatore)
- estraiamo un numero casuale uniforme $u \in [a, b]$ ed un altro numero casuale uniforme $v \in [c, d]$
- se $v < f(u)$, allora incrementiamo il contatore: $j = j + 1$
- iteriamo la procedura n volte

Allora, si ha

$$\int_a^b f(x)\mathrm{d}x \approx \frac{As}{n} \, .$$

In pratica, il metodo Monte Carlo conta quanti punti, di coordinate aleatorie (u, v), cadono sopra o sotto la curva di equazione f. Il numero di punti al di sotto della curva è s/n. Dunque, l'integrale che vogliamo calcolare (area della parte di piano) è $A \cdot \frac{s}{n}$.

In particolare, se scegliamo $f(x) = 1 - x^2$, $x \in [-1, 1]$, ci riduciamo a calcolare l'area del semicerchio di raggio 1, inscritto in un quadrato di lato 1. Con la tecnica appena descritta, ottentiamo che l'area cercata è π.

Esercizio 2.1. *Una compagnia di assicurazione ha* 25000 *polizze attive. Ciascun assicurato percepisce un risarcimento annuo che rappresenta una v.a. che si distribuisce con media pari a* 320 € *e scarto quadratico medio pari a* 540 €*. Qual è la probabilità che la compagnia paghi complessivamente* 8300000 €*?*

Esercizio 2.2. *Il numero ideale di studenti di un corso del primo anno di università è* 150*. Il management didattico dell'università sa che, in base agli anni precedenti, solo il* 30% *degli iscritti frequenta effettivamente i corsi, dunque decide di accettare fino a* 450 *nuove iscrizioni. Qual è la probabilità che il numero di studenti frequentanti sia superiore a* 150*?*